Neither separate nor equivalent

RUNNING HEAD: FEATURES IN OBJECTS

Neither separate nor equivalent: Relationships between feature representations within bound objects

Candice C. Morey

University of Groningen

Katherine Guérard

Université de Moncton

Sébastien Tremblay

Université Laval

Address correspondence to Candice C. Morey, Experimental Psychology, Grote Kruisstraat 2/1, 9712TS Groningen, The Netherlands. Email: c.c.morey@rug.nl

Please cite: Morey, C., Guerard, K., & Tremblay, S. (2013). Neither separate nor equivalent: Relationships between feature representations within bound objects. [References]. *Acta Psychologica*, *144*(2), 279–290. doi:http://dx.doi.org/10.1016/j.actpsy.2013.06.016

Word count: 9,854

Date submitted: 13 March 2013

Date of resubmission: 18 June 2013

Abstract

Evidence suggests that binding, or encoding a feature with respect to other features in time and space, can convey cognitive advantages. However, evidence across many kinds of stimuli and paradigms presents a mixed picture, alternatively showing cognitive costs or cognitive advantages associated with maintaining bound representations. We examined memory for colored letters drawn from similar and distinct color sets under circumstances that encouraged or discouraged the maintenance of color-letter binding. Our results confirmed previous change recognition research showing feature recognition improvement under explicit instructions to maintain binding. Color memory improved during binding, showing a reduced detrimental effect of feature similarity on retrieval, particularly when the letter served as the retrieval cue for a letter-color object. We found that feature recognition improved when two conditions were met: 1) relationships between features were to-be-remembered, and 2) the feature conjunction was relevant at test. Our results further suggest that this feature advantage arises because the encoded relationship between the features facilitates retrieval, not because features and objects are represented simultaneously in separate buffers.

Word count: 170

Keywords: working memory; binding; retrieval; short-term memory

Neither separate nor equivalent: Relationships between feature representations in bound objects

1. Introduction

The *binding problem* refers to the difficulty of elegantly explaining how features of various types and qualities, which elicit activation in disparate neural regions, meld to produce the percept of a unified object containing all these qualities. A further puzzle arises when one attempts to explain how these percepts are then represented in memory, which is known to be liable to severe limits (e.g., Cowan, 2001). How do disparate features become associated in memory, and does this process convey cognitive advantages or induce cognitive costs?

Several extant theories approach the binding problem from various levels of analysis. For very recently-perceived stimuli and their organization, Kahneman, Treisman, and Gibbs (1992) suggest that online perception is assisted by the formation of object files, which they describe as temporary representations that link aspects of an object at one time to its characteristics at a later time. Some construct like an object file seems necessary to explain how some visual image is perceived as having the same identity despite appearing successively in different locations, or even with different semantic identities. For example, Kahneman et al. describe the crowd's gradual identification of a flying object as Superman; first it was labeled a bird and then a plane before correct identification occurred, but each of these labels clearly refers to the same object in the environment, which is presumably represented in some stable manner even as some of its attributes change. The principles of the object file hypothesis have been

applied to the formation of coherent multi-modal object identities and responses across time (e.g., Zmigrod, Spapé, & Hommel, 2009) as well as to the association of component features of an object (Moore, Stephens, & Hein, 2010). A sufficient condition for the initial loading of component features into an object file is their appearance at the same location at the same moment in time (van Dam & Hommel, 2010; Xu, 2002). According to Kahneman et al., prior knowledge of similar objects (or long-term memory for the same object token) has no bearing on the creation of an object file, although certainly recognition of the correspondence between an attended object and the same previously encountered object can occur. Importantly though, to Kahneman et al. the object file is a temporary structure that supports perceptual organization and search of memory for recently viewed objects, requiring no link to more permanent knowledge.

Object files could however form the basis for more stable, persistent associations of features. With their Type-Token model, Zimmer and Ecker (2010) suggest that perceptual object files can become memorial object tokens, which can ultimately be consolidated and retrieved whole. Object tokens are believed to include an object's intrinsic features (Ecker, Maybery, & Zimmer, 2013) and are thought to underlie feelings of familiarity in recognition memory (Ecker, Zimmer, & Groh-Bordin, 2007). The biological mechanism believed to generate object token encoding is the synchronized firing of cell assemblies coding for each feature (Tallon-Baudry & Bertrand, 1999), and interestingly, modeled limits in the number of distinct synchronized groupings that may be simultaneously distinguished correspond to observable behavioral limits in memory (Cowan, 2001; Raffone & Wolters, 2001). The supposition that repeated occurrences of these groupings should lead to more stable memorial representations is in line with the

concept of Hebbian learning (e.g., Hebb, 1961). Thus the concepts of object files and tokens lie at the junction between perception and memory, at the very heart of the debate over how external information comes to persist in the mind.

Also occupying this nexus of perception and memory is the concept of working memory, which refers to the collection of memory and attention processes involved in real-time thinking. In any specific moment, memories that are activated for use may be considered the contents of working memory. Naturally, a comprehensive theoretical framework describing working memory should be able to explain how relationships between features are maintained. For the multi-component model of working memory (Baddeley & Hitch, 1974; Baddeley, 1986), explaining the effects of long-term knowledge on short-term memory (such as chunking; Cowan, Chen, & Rouder, 2004) or results suggesting that binding can occur between novel combinations of verbal and visual-spatial features (Prabhakaran, Narayaran, Zhao, & Gabrieli, 2000) posed major problems. Regarding verbal chunking, the multi-component model described a working memory system that is distinct from long-term memory, with no explicit means for interaction between new and old memories. Regarding cross-domain binding, verbal and visual-spatial representations were purportedly maintained in separate buffers, with no proposed storage mechanism capable of bridging them. Therefore the multi-component model could not explain how verbal labels and visual imagery (such as a name and a face) could come to be associated, nor could it explain why short-term memory for materials that are part of some learned structure was consistently superior to short-term memory for novel or unrelated information.

To address these problems, Baddeley (2000) proposed a new module for the multi-component model, the episodic buffer, which provided a store capable of holding information retrieved from secondary memory or from either of the domain-specific storage buffers. Originally, the episodic buffer was believed to be accessible only via the central executive (consistent with assumptions that binding requires attention, e.g., feature integration theory, Treisman & Gelade, 1980). However, tests of this hypothesis interpreted within the framework of the multi-component model violated this assumption. Specifically, Allen, Baddeley, and Hitch (2006) found that performing concurrent processing tasks, which should reduce the availability of the central executive for engaging in binding, had no more impact on memory for binding between visual features than on memory for the visual features only, a result that has been replicated many times over (e.g., Allen, Hitch, Mate, & Baddeley, 2012; Delvenne, Cleeremans, & Laloyaux, 2010; Morey & Bieler, 2013). In an updated exposition of the episodic buffer model, Baddeley, Allen, and Hitch (2011) proposed that features represented in the domain-specific short-term stores are passed to the episodic buffer. Instead of supposing that attention acts in a top-down manner to select information for representation in the episodic buffer, they proposed a filtering mechanism that can be tuned toward features meeting specified criteria. An explicit assumption of the updated framework is that features are represented in two ways, as disassociated representations in their domain-specific buffers and as components of object files, which reside in the episodic buffer.

The episodic buffer construct is not in conflict with the proposition that object files are created during perception and become memorial object tokens. Within the episodic buffer, Baddeley et al. (2011) accept the premise that features could be organized into

object files. However, it is not obvious that the multi-component working memory framework is necessary to explain differences between memories for objects and features. The episodic buffer is part of a framework that complicates predictions about perception and memory; it is thus necessary to consider whether the extra complexity it brings is really essential.

A unique assumption about feature storage during binding made in Baddeley's updated model is that when participants endeavor to maintain the binding between verbal and visual features, those features are represented twice within the working memory system. This assumption distinguishes the multi-component working memory system from frameworks that consider the contents of working memory as the most highly-activated portions of secondary memory (e.g., Cowan, 2005; Oberauer, 2009), which do not explicitly propose separate buffers for short-term maintenance. Moreover, this unique assumption provides a potential explanation for the cognitive advantages sometimes observed when participants undertake a memory task involving cross-domain binding. Intent to maintain verbal-spatial binding has been shown to improve recognition memory performance for individual features compared to a task that does not require binding (Morey, 2011), and also to change how domain-specific properties of an interfering task affect various memoranda. Morey (2009) tested memory for visually-presented letters and spatial locations in two groups of participants, one in which memory for the letter or location features (but never their binding) was always tested, and one in which binding was always tested, but in a manner that allowed inference about component feature memory. Both groups completed half of their trials with concurrent articulatory suppression. Unsurprisingly, memory for letters was always impaired by articulatory

suppression. However, participants in the binding group were better able to recognize letter features during suppression than participants in the feature group, suggesting that something about encoding letter-location associations helped to preserve letter representations from interference from articulation. One way to explain this result is to suppose that during the binding task letters were stored both in the phonological loop and as components of object files in the episodic buffer, and those maintained in the episodic buffer were shielded against interference from concurrent articulation.

However, there are other phenomena that seem inconsistent with the idea that features are stored in two forms during a binding task: under some circumstances cross-domain binding can also induce cognitive costs. Morey (2009) also found that even though memory for spatial locations in the feature group was never impaired during suppression, memory for spatial locations in the binding group was impaired by articulatory suppression (see also Kessels & Postma, 2002). Similarly, Guérard, Tremblay, and Saint-Aubin (2009) found that memory for sequences of spatial locations was impaired if the spatial locations were marked by phonologically similar letters compared to phonologically distinct letters. Guérard, Morey, Lagacé and Tremblay (2013) recently confirmed that while phonological similarity in a serial letter-location list impairs spatial memory, manipulating spatial complexity does not affect memory for the letters. These outcomes show that during binding, visual-spatial features can become vulnerable to sources of distracting information that typically exert selective interference on verbal memories. These outcomes are consistent with the proposition that during binding, features are stored in connection with each other, and that these connections may reduce within-domain feature interference while increasing cross-domain interference. However,

it is not clear that these outcomes are consistent with the idea that visual-spatial features are stored twice during a memory task that requires verbal-spatial binding in the manner described by Baddeley et al. (2011).

To attempt to resolve these contradictions, we aimed to test the effects of maintaining cross-domain binding on verbal and visual feature storage. We first set out to replicate previous results suggesting that binding can improve feature recognition memory, generalizing these effects to a context in which spatial location was not a to-be-remembered element. In Experiments 1 and 2, we used change detection tasks to measure recognition memory for colored letters, with conditions that enabled comparisons of feature memory during binding and feature memory when binding was not task-relevant. In each experiment, we tested recognition memory after a retention interval of a few seconds, so as to minimize the possibility that sensory memory representations remained active during test. We chose study and retention durations to be within ranges typically used in verbal short-term or working memory tasks (so that sub-vocalization of the letters would be possible), rather than durations of many minutes typically used in longer-term recognition memory tasks. We varied the similarity of the to-be-remembered color features so that we could test whether binding colors to letters helps to preserve color recognition from the expected detrimental effects of discriminating visually-similar stimuli during memory retrieval (Awh, Barton, & Vogel, 2007; Jalbert, Saint-Aubin, & Tremblay, 2008; Olsson & Poom, 2005). If features are represented twice within working memory during binding tasks, both as domain-specific features and components of cross-domain objects, then one would expect to observe 1) superior feature recognition

performance during tasks requiring maintenance of binding, as well as 2) a reduction in domain-specific interference effects (such as visual similarity effects) during binding.

Experiment 1 was designed to investigate both of these outcomes, making use of experimental blocks across which the type of test changed. We manipulated whether memory was tested with single-feature or bound probes that were embedded in blocks of pure or mixed trials. Pure blocks included 100% single-feature probes or 100% bound probes. In the mixed block, bound probes occurred on 75% of trials and single-feature probes in 25% of trials; participants therefore would have encoded the display with the expectation that a bound probe would probably follow. If the feature-level advantage is conferred during encoding by the expectation of a bound probe as predicted by the episodic buffer model of Baddeley et al. (2011), then recognition accuracy for single-feature probes in the mixed block – where binding is expected to occur – should be higher than those in the pure block. However, it could be that the cognitive advantages of binding are conferred during retrieval; if so, then feature recognition should be higher for bound probes, when participants make a judgment in reference to the entire multi-feature object, compared to single-feature probes.

2. Experiment 1

2.1 Method

2.1.1 Participants. Thirty-nine students of psychology at the University of Groningen completed this two-session experiment for partial credit towards a course requirement. One of these participants failed to meet our criterion for color vision, and two others performed at near-chance levels in at least one condition at set size 3. These

three participants were excluded from all analyses, leaving *N*=36 (9 males, 27 females) between 18 and 29 years old (*M*=20.44, *SD*=2.06).

2.1.2 Materials. Stimulus presentation and response collection were controlled with E-Prime 2.0 (Schneider, Eschmann, & Zuccolotto, 2002), using the keyboard. Stimuli were displayed on 17-inch CRT monitors set to resolution of 1024 x 768 pixels, with 100 Hz refresh rate and color temperature of 10,000 K. Each participant completed a short version of the Ishihara test (1966) to assess color vision before proceeding to the experimental task. This screening included 6 test items with single-digit responses.

Letters were selected randomly without replacement from the set G, H, J, M, Q, R, and Y. These 7 letters were chosen because their graphemes look different in upper- and lowercase forms and because they are phonologically distinct in English, Dutch, and German, the languages most commonly spoken by our likely participants. To visually distinguish letters and encourage phonological encoding of them, letters were presented in upper-case 18-point Palatino Linotype font (occupying 2.39 degrees of visual angle) at study and lower case 24-point Courier New font at test (occupying 4.35 degrees of visual angle).

A set of seven similar colors was chosen by selecting end points from neighboring hues (i.e., blue and yellow) and finding approximately equidistant points between them in 3-dimensional color space using the R-package *colorspace* (Ihaka, Murrel, Hornik, & Zeileis, 2012). Initially, we created several blue-yellow sets, and then examined these colors on the monitors in our laboratory to choose the set for which the closest neighbors seemed most discriminable. In our chosen set, the colors were all distinguishable from each other by the experimenters developing the stimuli when any two were presented

simultaneously. We excluded the colors used as end points from the final set, in order to increase the similarity between the colors. A set of seven distinct colors was also created. Both color sets are shown in Figure 1 along with their RGB coordinates.

2.1.3 Procedure. Participants were situated at personal computers in private booths. After giving written informed consent and completing the colorblindness screening, participants read instructions for the experimental task and then explained them to the experimenter, who either confirmed that the participant understood or corrected any misconceptions. Participants completed two sessions of three separate blocks including only single-feature probes, only bound probes, or a mix of single-feature and bound probes. Each session began with either a pure single-feature or a pure bound probe block. To ensure that participants understood instructions (and how they differed in each block), each participant was required to reach an accuracy criterion during the practice session, which included only displays of 3 items. If a participant did not respond correctly on at least 5 out of 8 trials, the practice session was repeated until criterion was reached.

After the first block of trials, participants were advised to take a break, and then given instructions for the mixed block of trials, which included both single-feature and bound probes. Instructions emphasized how the newly-added probe type differed from the tests experienced in the first block. Participants completed another practice session, with an accuracy criterion of 10 out of 16 trials needed to advance. Halfway through the mixed block, participants took a mandatory break of at least one minute. Finally, participants completed a final pure block of trials including either single-feature or bound probes (whichever condition they did not complete in the first block). In their second

session, participants completed this sequence with block order reversed. Experimenter errors in entering these order variables caused six participants to run the same order in both sessions, with three running each of the possible orders twice. An initial analysis showed no obvious effect of block order, so all of these participants were included in the reported analyses.

Participants initiated each trial with a button press. First, a fixation ("+") appeared centrally onscreen for 1000 ms, followed by a study array of 3 or 5 colored letters, each presented in one of five approximately equi-distant predetermined locations around an imaginary ellipse centered on the screen. Each location was separated by at least 5.45 degrees of visual angle. The study array remained onscreen 500 ms for each object, thus 1500 ms for 3 colored letters and 2500 ms for 5 colored letters, which approximately equated encoding time allowed for each object. After a 3000-ms blank retention interval, a probe appeared in the center of the screen. Trial events are depicted in section A of Figure 2.

Probe types are presented in sections B and C of Figure 2. In a letter single-feature probe, a letter was presented in white in the center of the screen. In a color single-feature probe, a colored # symbol was presented in the center of the screen. Participants were to press the "y" key (indicating "Yes") if the exemplar was presented in the study array and the "n" key (indicating "No") if it was not. The exemplar was a studied feature on 50% of trials.

For trials testing memory for binding, stimulus presentation was the same as in the single-feature trials, but at test, participants observed a colored letter in the center of the screen. Participants were instructed to press "y" if the colored letter matched one of

the colored letters presented at study, and otherwise to press “n”. On half of the trials, participants saw a target, in which the letter and its color were presented in the same combination as during study, which should elicit a “Yes” response. The remaining 50% were lures (which should elicit “No” responses). Three types of lures occurred with equal frequency: 1) a recombination lure included a color and a letter that were both presented at study, but within different objects; 2) a letter lure included a letter that was not presented at study in a color that was; and 3) a color lure included a color that was not presented at study on a letter that was presented.

Because we intended to analyze color and letter lure trials separately, we collected a large number of trials per participant. Each participant completed 240 trials within the pure bound-probe blocks, half of which tested lures, yielding 40 trials of each lure type. Each participant also completed 512 trials within the mixed blocks: 384 bound probes (192 targets and 64 each recombination, color and letter lures) and 128 single-feature probes (64 targets and 64 lures, half of each testing color and letter recognition). Additionally, each participant completed 192 trials in the pure single-feature blocks, yielding another 96 feature lure trials, 48 each of color and letter lure probes. Each session lasted approximately 90 minutes.

2.2 Results

It is apparent from the descriptive statistics given in Tables 1 and 2 and Figure 3 that recognition of letters as measured by the single-feature probes and letter lures within the bound probes was far better than recognition of colors. Separate analysis of these features must therefore be undertaken to avoid the report of theoretically uninteresting interactions involving a feature type variable. We estimated our analyses of variance

(ANOVA) models using the R package *BayesFactor* (version 0.9.0; R. Morey & Rouder, in preparation). This package calculates Bayes factors using the methods published by Rouder, Morey, Speckman, and Province (2012), yielding Bayes factors for combinations of effects entered by the user.

Interpretation of Bayes factors is straightforward: the more evidence in the data for a particular model, the higher the Bayes factor values compared to the model including only between-participant variance will be. Bayes factors greater than 1 indicate positive evidence, whereas Bayes factors between 0 and 1 indicate negative evidence. An advantage of Bayes factors for interpretation is that unlike *p* values, they can be directly compared with each other to weigh evidence. The output from the Bayes factor package yields Bayes factors for every possible combination of effects in an ANOVA model; though the best model yields the highest number, one can assess the strength of individual effects by comparing the winning model with models that exclude some factor or include additional factors, giving an idea of how strongly the evidence favors the inclusion of some factor in the model. Bayes factors can be expressed as favoring the null or the alternative hypothesis.

Although we report only Bayesian analyses, we also conducted traditional analyses, and the inferential outcomes generally agreed across these techniques. Where slight differences were observed, the Bayes factor analyses were more conservative; occasionally a factor that reached the $p<.05$ threshold for statistical significance in a traditional analysis was not present in the best model according to the Bayesian analysis[1].

[1] This was not the case for the null effects we interpret in Experiment 1; for instance, the effect of context block on single-feature probes was null by either analysis method, but the Bayes factor analysis allows us to interpret it whereas the traditional method does not.

However, Bayes factor analyses should be preferred over traditional methods because they provide a clearer means for interpreting effects (or lack of effects) than the *p* values given by a traditional analysis. Unlike the criterion logic of *p*-values, which allows interpretation of an effect as long as the *p*-value is ever so slightly less than some threshold, Bayes factors allow for the quantitative comparison of any two factors. For any analysis, we can report the extent to which including or excluding a factor (or an interaction of factors) improved the fit of a model.

All Bayes factor ANOVAs were estimated with 100,000 Monte Carlo iterations and were performed on proportions correct subjected to the arcsine square root transformation to insure adequate homogeneity of variance across conditions and groups. Mean proportions correct (with standard deviations) broken down by context block, trial type, color similarity, and set size can be found for the single-feature and bound probes in Tables 1 and 2, respectively. In section 2.2.1, we first compared mean proportions correct on the bound probe trials between the pure and mixed block in order to ensure that binding occurred in the mixed block. We then examined whether *encoding* binding modulates feature recognition by comparing letter and color recognition for single-feature trials between the mixed and pure blocks (section 2.2.2). Finally, to examine whether the advantage for feature recognition occurs during the *retrieval* of bound representations, we compared color and letter lure recognition between single-feature and bound probes (section 2.2.3). Regardless of probe or block type, the color and letter lure trials were alike in that complete memory for one kind of feature sufficed for a correct rejection. For instance, knowing all the letters presented, regardless of their colors, was enough to correctly reject a letter lure in a bound as well as in a single-feature probe. Because of

this, we could directly compare color and letter lure recognition using both single-feature and bound probes between context blocks to assess how memory for features changes as emphasis on encoding binding or retrieving binding increases.

2.2.1 Testing the assumption that binding occurred during the mixed blocks. In order to make inferences based on the assumption that participants maintained bound object representations during the mixed blocks of trials, it is necessary to confirm that performance on bound probes was comparable when the bound probes occurred in the mixed blocks compared to the bound blocks. If participants were not maintaining binding during the mixed blocks, which included some single-feature probes, then presumably their judgment of bound probes would be worse than in the pure bound block. We entered an ANOVA model on proportions correct (targets and all lure types included) on bound probe trials including factors of context (mixed or pure bound block), color similarity (distinct or similar), and set size (3 or 5). According to the Bayes factor ANOVA, the best model included only main effects of color similarity and set size (BF=2.17x10^{23} ± 1.32%[2]). To quantify the evidence *against* including an effect of context block, we computed the Bayes factor on the difference between the best model and the best model including an effect of block, which indicated that the simpler model was favored by a factor of about 12 (± 2.12%). Because there was substantial evidence that context block did not affect bound probe recognition, we can assume that participants were likely maintaining color-letter bindings similarly in both the pure and mixed block contexts. We

[2] For each analysis, we report the Bayes factor of the winning model against the model including only an effect of participant identity. This number reflects the extent to which our manipulations produced any effect, which is why it is typically so enormous. This number itself is less important for interpretation than the comparisons between the winning model and alternative models.

can thus proceed with examining whether color and letter memory (as measured by color and letter single-feature tests) improved during the mixed block context compared to the pure, single-feature context.

2.2.2 Does encoding binding improve feature memory? We analyzed proportions correct (including both target and lure trials) for color and letter single-feature probes in separate ANOVAs including context block (pure versus mixed), color similarity, and set size as factors. For color tests, the best model included only main effects of color similarity and set size (BF=1.05x10^{20}, ± 4.83%), and was preferred to the best model that included an effect of context block by a factor of 10.13 (± 7.59%). Recognition was better when color-probes came from distinct (M=0.79, SD=0.20) rather than similar (M=0.61, SD=0.26) color sets, and also when three (M=0.74, SD=0.23) versus five (M=0.66, SD=0.26) items were to be remembered.

For letters, the best model included only a main effect of set size (BF=45x10^{6}, ± 0.69%), and was favored over the best model that also included an effect of context block by a factor of 2.28 (±1.11%). Accuracy was better with three (M=0.93, SD=0.12) than five (M=0.87, SD=0.15) to-be-remembered items. The model including only set size was favored over models also including similarity by a factor of at least 2.59 (± 1.11%). Although the Bayes factor differences between models were not high, the size of the mean differences between distinct and similar colors or mixed and single-feature blocks (difference=0.02 in both cases) does not suggest any theoretically interesting effect. Furthermore, performance on letter tests during mixed blocks was if anything slightly worse than performance during pure blocks; if features are stored twice when binding is occurring, one would have expected improved feature recognition during binding.

In summary, performance on neither color nor letter single-feature tests improved in the mixed block context compared to the pure single-feature context, suggesting that any advantage for feature memory that may occur during binding (such as that previously observed by Kessels & Postma, 2002; Morey, 2009) was not due merely to the encoding of features in multiple formats during binding.

2.2.3 Does retrieval of bound representations improve feature memory? We compared recognition of lures when they were presented within single-feature or bound probes to examine how retrieving bound representations affected feature memory. Separately for both color and letter lures, we carried out 3-way ANOVAs with probe type (single-feature or bound), color similarity, and set size as factors. For color lures, the best model included main effects of probe type, color similarity, and set size, plus an interaction between probe type and color similarity (BF=2.56x10^{68}, ± 1.47%). This analysis is depicted by the warm-colored lines in Figure 3. This interaction was due to a reduction in the difference between distinct and similar colors with bound probes compared with single-feature probes. Compared to the best model excluding the interaction between test type and color similarity, the winning model including this interaction was favored by a factor of 33.64 (± 1.81%).

For completeness, we also carried out the same analysis on letter lures (refer to the cool-colored lines in Figure 3). Considering that mean differences in letter recognition were quite small, these results should be treated cautiously. The best model to emerge from a Bayes factor ANOVA included only main effects of color similarity and set size (BF=65.14x10^{5}, ± 0.08%). While the effect of set size seems clear, the effect of color similarity appears much smaller, so we compared the model including similarity

and set size with a model including only set size. The model including an effect of color similarity was only slightly favored over the model excluding the effect (BF=1.11 ± 1.03%).

2.3 Discussion

These results suggest specifically that maintaining and retrieving visual features as components of bound object representations changes the quality of visual feature memory. In this case, we showed that recognition of color bound with a letter identity improved color memory compared with circumstances in which colors could be maintained separately from letters. Retrieving binding protected color feature memory from some of the impairment expected when several similar colors were to be maintained simultaneously. An improvement in color memory during binding could be consistent with the improvement observed in letter memory during articulatory suppression when letters are retained as bound representations (Morey, 2009).

The design of Experiment 1, which included blocks that mixed single-feature and bound probes, allowed a test of the hypothesis that features are stored in multiple formats during a binding task (Baddeley, et al., 2011) but otherwise are only stored in a domain-specific format. Bayes factor ANOVAs allowed for the quantification of evidence favoring the null hypothesis that accuracy on single-feature probes was equivalent during pure and mixed block contexts. These analyses suggested that encouraging encoding of binding did not affect feature memory as measured with single-feature tests. These findings cast doubt on the notion that features are stored both as separate entities and as components of object files during binding, suggesting that we must seek another explanation for the reduced effects of domain-specific impairments that have been observed during binding tasks (e.g., Kessels & Postma, 2002; Morey, 2009).

It has been shown that factors that selectively impair verbal memory, such as phonological similarity (Guérard et al., 2009) or articulatory suppression (Kessels & Postma, 2002; Morey, 2009) carry over to non-verbal features during binding. We did not find strong evidence that this occurs in the opposite direction, namely from non-verbal to verbal features. Color similarity slightly affected letter feature recognition, but this effect did not interact with probe type, so there is no reason to suppose that this effect occurred because of processes specifically involved in maintaining binding. However, we do not think that we can make any strong claim about the absence of this interaction, or indeed about letter feature memory generally from these data. Letter feature memory was near ceiling, while color feature memory was not. With paradigms more conducive to matching verbal and visual-spatial recognition performance, we have shown that spatial complexity manipulations do not spill-over to verbal memory (Guérard, et al., 2013), so we did not necessarily expect to observe any effect of color similarity on letter feature memory.

With the following experiments, we try to better understand why this improvement to feature memory during retrieval of binding arises. One possible explanation could be that the mere presence of the familiar letter facilitated retrieval for the color associated with it at study. Perhaps this effect could be due to obligatory binding that occurs irrespective of a task requirement to retain associations between features. If our bound stimulus presentation format provoked obligatory encoding of letter-color binding, then of course the differences between block contexts we depended on to test the dual feature-storage hypothesis would be obscured. Before we can consider how retrieving bound objects produces this advantage, it is first necessary to show that

this effect does not occur whenever both features are available at test, regardless of whether retrieving a bound object is required.

With Experiment 2, we repeated critical aspects of the single-feature memory tests of Experiment 1, except that at test the irrelevant (but potentially helpful) other feature type was also present and thus able to serve as a retrieval cue. If the benefit we observed in the bound probes of Experiment 1 occurred solely because of the presence of the other feature at test, we will observe the same benefits in Experiment 2. With this experimental design, we can compare accuracy and latency towards intact and recombined letter-color pairings to implicitly assess whether bindings from study were remembered (cf. Elsley & Parmentier, 2009; Prabhakaran, Narayaran, Zhao, & Gabrieli, 2000).

3. Experiment 2

3.1 Method

3.1.1 Participants. Twenty-nine psychology students from the University of Groningen participated in this study for their choice of credit towards a course requirement or a €7 honorarium. One participant's data were excluded from analyses due to imperfect performance on a color-blindness screening, yielding a final N=28 (17 females and 11 males), from 18 to 27 years old (M=22.14, SD=2.45).

3.1.2 Materials. Apparatus and stimuli were the same as in Experiment 1.

3.1.3 Procedure. The procedure was similar to that of Experiment 1. After indicating consent to take part in the study and completing a short color-blindness screening, participants received instructions and completed a supervised practice session of 10 trials with accuracy feedback. Because we planned to analyze response times as an

indicator of obligatory binding, we provided accuracy feedback throughout the experimental session and instructed participants to respond as quickly as possible while avoiding errors.

Events in each trial were similar to Experiment 1, except that we increased study time slightly (to 600 ms per object, thus 1800 ms for 3 and 3100 ms for 5 colored letters) and the tests were always based on a single feature. At test, participants saw a colored letter with a prompt of either "Letter?", indicating that they should base their response on whether they remembered the letter identity from the study array, or "Color?", indicating that they should base their response on their memory for the color only. Unlike the single-feature probes of Experiment 1, in which participants judged one feature without an irrelevant feature, participants in Experiment 2 always judged one feature with an untested feature also present. Half of the trials were targets, where the tested feature was presented at study and half were lures, where the tested feature was not presented at study. In half of the target trials, the colored letter was an exact match of a studied color-letter combination and in the remaining target trials the letter was the color of a different studied letter. For lures, the untested feature was always a studied feature, as in the feature lure trials within bound probes in Experiment 1. A total of 192 experimental trials were collected from each participant in a 45-minute session.

3.2 Results

Tables 3 and 4 give the mean proportions correct and response times (and their standard deviations) for letter and color feature tests respectively, broken down by test type (matching targets, recombined targets, and lures), color set (similar and distinct), and set size (3 and 5). All analyses of proportion correct were subjected to the arcsine square

root transformation. We analyzed target and lure trials separately. In our analysis of target trials, our aim was to look for evidence that obligatory binding occurred by comparing matching and recombined targets; if binding occurred, then clear advantages for matching targets should be observed in both accuracy and response latency, even though one feature dimension was always irrelevant at test. We then compared lure recognition in Experiment 2 with lure recognition for the single-feature and bound probes of Experiment 1.

3.2.1 Evidence of obligatory binding in target accuracy? We ran 3-way ANOVAs of proportions correct on letter targets with probe type (matching target, recombined target), color set (similar, distinct), and set size (3,5) as factors, and the best model of these data included only a main effect of set size (BF=305 ± 0.71%). Comparisons of Bayes factors with competing models indicated that this single-effect model fit the data better than a model that also included a main effect of probe type by a factor of 6 (± 1.15%) and fit better than a model that included an interaction between probe type and set size by a factor of 4 (± 1.26%).

We ran the same 3-way ANOVA on color target accuracies. The best model emerging from this analysis included only an effect of color similarity (BF=0.61± 1.66%). However, this Bayes factor of <1 indicated that the model including only between-participants variance is favored. There is thus no evidence in these data of an advantage with matching compared to recombined targets.

3.2.2 Evidence of obligatory binding in target latency? The same analyses were run on median response times, excluding inaccurate responses and responses faster than 180 ms. For letter targets, the best model included only an effect of set size (BF=320

± 0.68%). This model, which excluded probe type, only favored a model also including an effect of probe type by a factor of 1.84 (± 1.08%), but if anything, responses to the matching targets (*M*=1859) were somewhat slower than responses to the recombined targets (*M*=1793), not faster as one would expect if the match between study and test conditions facilitated retrieval. For color targets, no model produced a Bayes factor greater than 1; the best model included a main effect of set size (*BF*=0.17 ± 0.67%). This means that a model including only between-participant variance is preferred over models including any combination of the specified independent variables.

3.2.3 Comparison of lure recognition across Experiments 1 and 2. If binding occurred in this task and memory for binding were accessible after the 3000-ms retention interval, one would expect to see clear advantages for matching targets over recombined targets in accuracy, latency, or both dependent measures. Neither accuracy nor latency analyses of targets showed this. We therefore assume that obligatory binding, at least strong enough for the representation to persist throughout our 3000-ms retention interval, did not occur in our paradigm. Assuming that binding occurred in the bound probe conditions of Experiment 1 but not in Experiment 2 where it was not task-relevant, we set out to test whether the effect of color similarity varied as a function of binding by comparing lure recognition between Experiment 1 and Experiment 2. We entered color lure recognition accuracy into 3-way Bayes factor ANOVAs with group (Experiment 1 or 2) as a between-participants factor, and color similarity and set size as within-participants factors. We compared color lure recognition in Experiment 2 with bound probe color lure recognition of Experiment 1, and then separately with single-feature lure recognition of Experiment 1. These cross-experiment analyses turned up different best-fitting models.

An effect of group (as well as interactions between group and similarity and group and set size, and main effects of similarity and set size) was present in the best model comparing trials from the bound probes of Experiment 1 with Experiment 2 (BF=3.69x10^{54}, ± 1.64%), but when comparing the single-feature probes of Experiment 1 with Experiment 2, neither a main effect of group nor any interactions involving group were present in the best model, which included only main effects of similarity and set size (BF=3.24x10^{56}, ± 1.14%). This result suggests that feature recognition in Experiment 2 (and the effects of color similarity on it) was more comparable to that of the single-feature than the bound probes of Experiment 1. For completeness we also compared letter lure recognition across experiments, but found no differences between the results of Experiment 1 and those of Experiment 2.

3.3 Discussion

The results of Experiment 2 eliminated the possibility that the protection afforded to color representations during binding could be due to obligatory encoding of binding occurring whenever bound color-letters were presented. Evaluating a bound probe in Experiment 1 led to an improvement in similar color lure recognition, but did not in Experiment 2, where memory for binding was not necessary to perform the recognition task. Thus the mere presence of the other feature at test does not convey a binding advantage. Furthermore, comparing accuracy and latency to matched and mismatched targets yielded no evidence that color-letter bindings were retained in Experiment 2, where they were unnecessary at test.

Taken together, Experiments 1 and 2 show improvements in feature memory under circumstances in which both encoding binding is encouraged *and* retrieving binding is relevant at test. Therefore, this improvement seems to be attributable to the

retrieval of bound representations rather than to the requirement to bind features during encoding, or even to obligatory binding during encoding. However, the change detection task used in Experiments 1 and 2 might not be the optimal paradigm for measuring memory for binding, because we can only discern whether the participant believes that the item was present or absent; we do not know whether the participant judged based on memory for features only or memory for binding. In order to offer an explanation for the protection from within-domain interference apparently afforded during retrieval of binding, we conducted a final experiment in which we probed recognition of one feature via its companion, a more direct test of memory for binding. As in Experiments 1 and 2, participants were given a display of colored letters to remember, but instead of indicating memory through recognition of an exemplar probe, participants were shown either a studied letter or a studied color at test and were asked to choose the feature that was bound with the probed feature at study from an array of choices. This design affords an examination of how features within bound representations are retrieved and the potential mechanism responsible for the beneficial effect of binding on feature memory. Bound objects offer two features that can be used as retrieval cues at test. Both features might not be equally vulnerable to the detrimental effect of color similarity. These data will afford new insights into the reasons underlying the unique advantages we observed with bound probes in change detection.

4. Experiment 3

4.1 Method

4.1.1 Participants. Thirty-one volunteers participated in this study for their choice of credit toward a course requirement or an honorarium of €7. Four participants'

data were excluded from all analyses because of imperfect performance on a 6-item colorblindness assessment (Ishihara, 1966). The final sample included 27 participants (7 males) ranging from 19 to 27 years old (M=21.52, SD=1.81).

4.1.2 Materials. Materials were similar to those used in previous experiments. Letters and similar and dissimilar colors came from the same sets described in Experiment 1.

4.1.3 Procedure. Participants were situated at personal computers in private booths. After giving written informed consent and completing a colorblindness screening, participants read instructions for the experimental task and then explained them to the experimenter, who either confirmed that the participant understood or corrected any misconceptions.

Study arrays included colored letters and remained onscreen for 600 ms per object. Participants were usually tested on their memory for the binding between letter and color features. This could occur in either of two equally-likely ways. As shown in Figure 4, participants might see a single white letter and a list of randomly-ordered color choices. In this test, the objective was to choose the color that was presented with the given letter at study. Alternatively, participants could see a single color square with a list of randomly-ordered letter choices; in this case, participants should choose the letter that was presented with the given color at study. Seventy-five percent of the trials in this block ended with a binding test. In the remaining trials, participants were tested on their memory for either the studied letter or color features. In these tests, participants were shown either an array of all the possible letters that could occur or all the possible colors that could occur, and were instructed to choose the 3 or 5 features that they observed

during the trial. When the participant selected an item from the choice array, that item moved to a response set in the lower portion of the screen, and could not be selected again. Feature recall tests were equally likely to query letters or colors. Stimulus presentation was the same for all trials, so that participants could not predict whether a binding or a feature recall test was coming. Each participant completed 96 binding test trials and 32 feature recall trials in total.

4.2 Results

We were primarily interested in analyzing the binding test trials, in order to compare effects of color similarity on recognition accuracy when the color versus the letter served as the eliciting probe feature. For completeness, mean accuracy (and standard deviations) on the feature recall trials are also given in Table 5. These descriptive statistics make clear that feature recall for these displays was quite high overall, and nearly exhaustive for letter features.

In analyzing binding tests, we considered two different categories of responses. Out of the seven possible choices, the participant might select an exemplar that was actually present at study (i.e. the correct feature or a studied feature that was not bound to the given feature; a *recombination error*) or one that was not (an *intrusion error*). We analyzed these responses separately, because we considered that intrusion errors were more likely to reflect guessing due to incomplete feature memory than poor binding memory. Given the rates of feature recall reported in Table 5, participants were more likely to exhaustively recall the letter than the color features.

4.2.1 Intrusion error analysis. As one might expect based on the strong feature-only memory reported in Table 5, intrusion errors were rare. We therefore collapsed

across set sizes to prevent an unbalanced design in which some subjects infrequently or never committed intrusion errors in some cells and ran a 2-way ANOVA with eliciting feature (color or letter) and color similarity as factors on proportions of intrusion errors. The best model that emerged from a Bayes factor ANOVA included effects of both eliciting feature and color similarity and an interaction between these factors (BF=4.85x10^{14}, ± 0.86%). Intrusions were more likely when letter (M=0.31, SD=0.16) rather than color (M=0.12, SD=0.14) was the eliciting feature; this matches the feature recall performance in Table 5, which shows letter recall near ceiling and lower performance on color feature recall. The effect of color similarity unsurprisingly depended upon test. This relationship is shown in the left panel of Figure 5; color similarity impacted color recall as expected.

4.2.2 Recombination error analysis. Excluding trials in which an intrusion error was made so that we could restrict our analysis as much as possible to binding recognition accuracy given feature memory, we ran ANOVAs with eliciting feature, color set, and set size on binding judgment accuracy. The best model included effects of eliciting feature, color similarity, and set size plus interactions between eliciting feature and color similarity and eliciting feature and set size (BF=3.50x10^{25}, ± 1.14%). The best model was favored over a model including the same effects but no interactions by a factor of 14.39 (± 1.43%). Accuracies were higher when color elicited a letter response (M=0.56, SD=0.25) than when a letter probe elicited a color response (M=0.46, SD=0.23), higher with distinct (M=0.54, SD=0.25) than similar (M=0.48, SD=0.22) colors, and higher with 3 (M=0.63, SD=0.21) than 5 to-be-remembered objects (M=0.39, SD=0.21). As Figure 5 makes clear, the interaction between eliciting feature and color set

occurred because color similarity impaired the choice of a letter given a color much more than it impaired the choice of a color given a letter.

Recognition was well above chance performance (i.e., approximately 0.14 with 7 choices), therefore the interaction between eliciting feature and color similarity is not likely restricted by floor effects. Given the analysis of intrusions reported above, we think that the main effect of eliciting feature reflects superior letter feature memory to some extent. If we assume that memory for letters was nearly perfect, then we can suppose that in the trials where a given color elicited a letter response, participants eliminated the unpresented letters as plausible choices. As a consequence, the probability of guessing correctly increased from 1/(7 choices) to 1/(set size; either 3 or 5). If we assume that proportion correct is a function of both successful retrieval and guessing, we should correct for guessing. Assuming a guessing rate of 1/(set size) for the color-choose-letter trials and a guessing rate of 1/7 for the letter-choose-color trials produces very similar estimates of retrieval success across conditions (for 3 items, $M_{\text{Letter-choose-Color}}$=0.42, compared to $M_{\text{Color-choose-Letter}}$=0.38; for 5 items, $M_{\text{Letter-choose-Color}}$=0.23, compared to $M_{\text{Color-choose-Letter}}$=0.21).

4.3 Discussion

With Experiment 3 we explored choice recognition accuracy as a function of the feature eliciting the judgment. Although accuracy was lower when a letter elicited the corresponding color, judgments in this condition were not affected by color similarity, whereas choices of a letter elicited by color were impaired by color similarity. Building upon Experiments 1 and 2, which suggested that advantages to feature memory from binding processes do not arise during encoding, this result suggests that binding might prevent the detrimental effect of color similarity via retrieval processes. When two

features are part of the same bound object, presumably either feature can serve as a retrieval cue for the other. Experiment 3's results suggest that when the letter (which is less affected by color similarity) served as the retrieval cue, the effect of color similarity on recall was diminished. In the General Discussion, we consider what this finding suggests for theories of feature binding in short- and long-term memory.

5. General Discussion

We examined memory for binding between randomly paired letters and colors, features that can be considered a simpler case of more complex cross-domain associations, such as the pairing between a name and a face. We started by measuring change recognition accuracy, comparing accuracy on probes of single-feature memory with accuracy on bound probes composed of feature lures, which could be correctly rejected with sufficient feature memory alone. In Experiment 1 we found that feature recognition did not improve when single-feature probes were mixed with bound probes, presumably creating a need to encode bindings. However, compared to single-feature probes, bound probes reduced the damaging effect of color similarity on color recognition. With Experiment 2, we ruled out the possibility that this advantage arose from obligatory encoding of letter-color binding. In Experiment 3, we used a choice recognition task to test memory for binding by asking participants to choose the color that matched a given letter, or alternatively the letter that matched a given color. Results showed that mitigation of the impairing effect of color similarity occurred only when the letter probed the color. Taken together, these findings do not support the idea that cognitive advantages afforded by binding arise because some features are stored simultaneously in multiple formats, organized by invoking unique processes during

encoding. Instead, the protection that binding affords to feature memory more likely occurs because of processes occurring during retrieval, and differs depending on how the feature is cued for retrieval.

These results are not favorable for the proposition raised by Baddeley et al. (2011), that features are encoded in working memory into separate, domain-specific feature stores, and then copied to bound object representations in the domain-general episodic buffer. This incarnation of the episodic buffer model makes the clear prediction that feature storage should be superior in contexts where the episodic buffer is needed to maintain object representations, which was not borne out by our data. This unique prediction could have explained how the cognitive advantages afforded during cross-domain binding arise (e.g., Kessels & Postma, 2002; Morey, 2009). Though Baddeley et al.'s hypothesis could explain the reduced effects of color similarity during binding that we observed, our finding of no comparable improvement for single-feature tests during a binding context falsifies the hypothesis that information encoded in domain-specific buffers is subsequently copied to the domain-general episodic buffer. We must therefore look elsewhere for a way to explain both of these findings simultaneously.

We confirmed that binding conveys protection to bound features against within-domain interference. If this advantage occurs only upon the retrieval of a bound object representation, and even in that case, only when a particular component feature of the object serves as the retrieval cue, how can it best be explained? Even for our paradigm, which measured memory for arbitrarily-colored letters after a brief retention interval, we think that this phenomenon is best explained in terms of the same processes used to explain long-term episodic memory. Given some cue, retrieval is believed to involve

sampling of representations in some way related to that cue, which are then searched and compared to the cue until a match is found (Search of Associative Memory (SAM); Raaijmakers & Shiffrin, 1981). The SAM model assumes that typical interference patterns can be explained as resulting simply from response competition, which occurs as sets of potential responses are generated. Variations of this model have successfully predicted several classic memory effects, including retroactive and proactive interference, and Osgood's (1949) retroactive surface transfer hypothesis (Mensink & Raaijmakers, 1988).

Osgood (1949) proposed a descriptive explanation of effects of similarity on stimuli and responses, the crux of which is the prediction of an asymmetric effect of similarity on retrieval depending on whether the stimuli eliciting retrieval or the potential responses were more similar. According to Osgood's principle, interference due to similarity of the stimuli, rather than of the responses, produces worse interference. A stimulus from a similar set serves as a poor retrieval cue because there are multiple potential sources for interference: in the event that the stimulus comes from a similar set, the participant might confuse the given stimulus with one of the other stimuli (and thus retrieve the wrong response), as well as forget which of the response choices was paired with the eliciting stimulus. However, when the stimuli are distinct from each other but the response choices are similar, confusing one stimulus for another becomes less plausible. Although Osgood's hypothesis was supported primarily with research on interference in verbal paired associate learning (see Anderson, Green, and McCulloch (2000) for a review), this hypothesis can also be applied to other stimuli and paradigms. Our results in Experiment 3 may be considered in light of this principle: a stimulus from a similar set

(i.e., similar colors) facilitated retrieval from a distinct set of responses (i.e., letters) less than stimuli from a distinct set (i.e., distinct colors), but when eliciting stimuli came from the always distinct and nearly perfectly recalled letter set, the similarity of the response choices was inconsequential.

Several existing models of working memory assume that working memory processes are embedded within the broader memory system (e.g., Cowan, 2005; Oberauer, 2009). Assuming such a framework would alleviate the need to differentiate between short-term and long-term memory for binding, because in either case, active representations might be subjected to the same processes. For instance, Cowan's framework supposes that working memory consists of representations active within long-term memory, with a small subset of these in a state of extra high activation known as the focus of attention. The capacity of the focus of attention could be as low as a single object (Oberauer, 2009), or it could vary across individuals over a small range (3± 1; Cowan, 2001). While retrieving an object from the focus of attention would not involve search of long-term memory, retrieving a recently activated object that is not currently in the focus of attention could be supposed to involve the same processes used to retrieve any long-term knowledge. This sort of framework involves fewer assumptions than Baddeley's multi-component model, which supposes that several categories of information can be simultaneously held in short-term stores separately from long-term memory.

The asymmetry we observed between features serving as retrieval cues in our letter-color choice recognition task of Experiment 3 can potentially explain the modest size of the mitigation of the color similarity effect we observed with our change detection

paradigm in Experiment 1. Although the presence of the critical interaction between test type and color similarity was robust, similarity strongly impaired color lure recognition even during binding. In our change detection bound probes, the recognition probe included both letter and color information, thus either feature could serve as the basis for retrieving object representations, possibly weakening the potential for alleviation of the similarity effect. However, in Experiment 3, the recognition probe included only one feature, and here we observed complete obliteration of the similarity effect with distinct letter stimuli as elicitors.

Another finding that should be discussed is the small effect of color similarity on letter lure recognition we observed in Experiment 1. This suggests that some interference between colors and letters occurred in our paradigm. In other work, we have observed that manipulations of path complexity (e.g., Parmentier, Elford, & Maybery, 2005) of verbal-spatial objects have no effect on recall of the associated verbal items (Guérard, et al., 2013). Possibly, our manipulation of spatial complexity was not comparable enough to visual or phonological similarity to produce the same kind of retrieval interference. Another possibility is that the type of memory task used is critical for determining whether the detrimental effect associated with one feature will spread to the other feature. For instance, in our previous paper (Guérard et al., 2013), we suggested that in a serial memory task, where reliance on verbal rehearsal is strong, letters could be the primary feature around which other features (e.g., spatial location) were encoded. The primary feature could then be retrieved without reference to the other features. In other paradigms such as the change detection task used in Experiment 1, letters might not serve the same function, leaving them vulnerable to detrimental effects associated with other features.

Here we cannot strongly claim that one feature dimension was the primary dimension for recall, but the finding that letter recall was slightly affected by color similarity suggests that letters were not recalled independently of the color. Though small, this effect is in line with multiple studies showing that letter features were affected by non-verbal manipulations in a change detection paradigm (e.g. Campo et al., 2010; Jiang, Olson, & Chun, 2000). As mentioned above, in our change detection paradigm, both color and letter were present at test and therefore either could have functioned as a retrieval cue.

5.1 Conclusions

In conclusion, our results suggest that a framework supposing the simultaneous storage of features both as individual representations and components of objects (like the multi-component model; Baddeley, et al., 2011) is more complex than necessary. Such a framework does not elegantly account for the range of experimental outcomes observed when memory for binding is tested. We found no evidence supporting the proposition of dual storage of visual color features. Rather, we promote the idea that near-immediate memory for binding is a function of precisely the same processes assumed in models of secondary memory: a probabilistic search of memory dependent on retrieval cues. The assumption that short- and long-term memories are subject to the same retrieval processes can be accommodated by models of working memory that assume that working memory is a subset of broader memory system (Cowan, 2005; Oberauer, 2009), but are not predicted by the current assumptions of the episodic buffer of the multi-component working memory model (Baddeley, et al., 2011).

Acknowledgments

We thank Indira Pliska, Anne-Meryem Grabski, Malte Bieler, Rieka Buggentien, Nelly Rahmede, Michael Wolff, and Lydia Beck for assistance with data collection.

References

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*, 298–313.

Allen, R. J., Hitch, G. J., Mate, J., & Baddeley, A. D. (2012). Feature binding and attention in working memory: A resolution of previous contradictory findings. *Quarterly Journal of Experimental Psychology, 65*, 2369-2383.

Anderson, M.C., Green, C., & McCulloch, K.C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1141-1159.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*, 1393–1400.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In *Recent Advances in Learning and Motivation Volume VIII*. New York, NY: Academic Press.

Campo, P., Poch, C., Parmentier, F., Moratti, S., Elsley, J., Castellanos, N., Ruiz-Vargas, J.M., del Pozo, F., & Maestu, F. (2010). Oscillatory activity in prefrontal and posterior regions during implicit letter-location binding. *Neuroimage*, *49*, 2807–2815.

Cousineau, D. (2005). Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson's method. *Tutorial for Quantitative Methods in Psychology*, *1*, 42–45.

Cowan, N. (2005). *Working Memory Capacity*. New York, NY: Psychology Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, *15*, 634–640.

Delvenne, J.F., Cleeremans, A., & Laloyaux, C. (2010). Feature bindings are maintained in visual short-term memory without sustained focused attention. *Experimental Psychology*, *57*, 108–116.

Ecker, U. K. H., Maybery, M., & Zimmer, H. D. (2013). Binding of intrinsic and extrinsic features in working memory. *Journal of Experimental Psychology: General, 142,* 218-234.

Ecker, U. K. H., Zimmer, H. D., & Groh-Bordin, C. (2007). Color and context: An ERP study on intrinsic and extrinsic feature binding in episodic memory. *Memory & Cognition*, *35*, 1483–1501.

Elsley, J. V., & Parmentier, F. B. R. (2009). Is verbal–spatial binding in working memory impaired by a concurrent memory load? *The Quarterly Journal of Experimental Psychology*, *62*, 1696–1705.

Guérard, K., Tremblay, S., & Saint-Aubin, J. (2009). Similarity and binding in memory: Bound to be detrimental. *The Quarterly Journal of Experimental Psychology*, *62*, 26–32.

Guérard, K.G., Morey, C.C., Lagacé, S., & Tremblay, S. (2013). Asymmetric binding in serial memory for verbal and spatial information. *Memory & Cognition, 41, 378-391*.

Hebb, D. O. (1961). *The organization of behavior: A neuropsychological theory*. Oxford England: Basic Books.

Ihaka, R., Murrel, P., Hornik, K., & Zeileis, A. (2012). *Package "colorspace"* (Vol. http://cran.r-project.org/web/packages/colorspace/index.html).

Ishihara, S. (1966). *Tests for colourblindness*. Tokyo, Japan: Kanehara Shuppan.

Jalbert, A., Saint-Aubin, J., & Tremblay, S. (2008). Visual similarity in short-term recall for where and when. *The Quarterly Journal of Experimental Psychology*, *61*, 353–360.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 683–702.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.

Kessels, R. P. C., & Postma, A. (2002). Verbal interference during encoding and maintenance of spatial information in working memory. *Current Psychology Letters: Behaviour, Brain & Cognition*, *9*, 39–46.

Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455.

Moore, C. M., Stephens, T., & Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin & Review*, *17*, 731–736.

Morey, C. C. (2011). Maintaining binding in working memory: Comparing the effects of intentional goals and incidental affordances. *Consciousness and Cognition: An International Journal*, *20*, 920–927.

Morey, C. C. (2009). Integrated cross-domain object storage in working memory: Evidence from a verbal-spatial memory task. *Quarterly Journal of Experimental Psychology*, *62*, 2235–2251.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial for Quantitative Methods in Psychology*, *4*, 61–64.

Morey, C.C., & Bieler, M. (2013). Visual working memory always requires general attention. *Psychonomic Bulletin & Review, 20*, 163-170.

Morey, R.D., & Rouder, J. N. (in preparation). Bayes factors for evidence-based inference in common psychological designs.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation: Advances in research and theory*, *51*, 45–100.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 8776–8780.

Osgood, C. E. (1949). The similarity paradox in human learning: a resolution. *Psychological Review*, *56*, 132–143.

Parmentier, F. B. R., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 412–427.

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience*, *3*, 85–90.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, *13*, 766–785.

Rouder, J. N., Morey, R.D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356-374.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*, 151–162.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Van Dam, W. O., & Hommel, B. (2010). How object-specific are object files? Evidence for integration by location. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1184-1192.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 458–468.

Zimmer, H. D., & Ecker, U. K. H. (2010). Remembering perceptual features unequally bound in object and episodic tokens: Neural mechanisms and their

electrophysiological correlates. *Neuroscience and Biobehavioral Reviews, 34,* 1066-1079.

Zmigrod, S., Spapé, M., & Hommel, B. (2009). Intermodal event files: Integrating features across vision, audition, taction, and action. *Psychological Research*, *73*, 674–684.

Table 1: Proportions correct (with standard deviations) on single-feature probes in the pure and mixed blocks, Experiment 1

Pure Feature Block (100% single-feature probes)

	Distinct		Similar	
	3	5	3	5
Letter Targets	0.93(0.11)	0.91(0.12)	0.93(0.09)	0.90(0.14)
Color Targets	0.78(0.18)	0.64(0.19)	0.73(0.20)	0.75(0.18)
Letter Lures	0.95(0.10)	0.88(0.14)	0.94(0.10)	0.86(0.18)
Color Lures	0.91(0.13)	0.86(0.16)	0.56(0.21)	0.37(0.24)

Mixed Block (25% single-feature probes)

	Distinct		Similar	
	3	5	3	5
Letter Targets	0.93(0.11)	0.90(0.11)	0.91(0.13)	0.89(0.13)
Color Targets	0.73(0.21)	0.66(0.19)	0.78(0.17)	0.78(0.20)
Letter Lures	0.94(0.11)	0.82(0.20)	0.89(0.17)	0.83(0.18)
Color Lures	0.91(0.13)	0.83(0.19)	0.52(0.24)	0.39(0.20)

Note. N=36.

Table 2: Proportions correct (with standard deviations) on bound probes in the pure and mixed blocks, Experiment 1

Pure Binding Block (100% bound probes)

	Distinct		Similar	
	3	5	3	5
Targets	0.81(0.15)	0.63(0.18)	0.78(0.14)	0.69(0.17)
Recombination Lures	0.66(0.24)	0.60(0.21)	0.56(0.25)	0.44(0.22)
Letter Lures	0.92(0.18)	0.86(0.18)	0.89(0.19)	0.83(0.21)
Color Lures	0.90(0.10)	0.81(0.16)	0.60(0.18)	0.47(0.27)

Mixed Block (75% bound probes)

	Distinct		Similar	
	3	5	3	5
Targets	0.81(0.14)	0.66(0.19)	0.79(0.12)	0.74(0.15)
Recombination Lures	0.66(0.24)	0.57(0.20)	0.50(0.21)	0.42(0.22)
Letter Lures	0.89(0.21)	0.87(0.19)	0.89(0.19)	0.84(0.18)
Color Lures	0.86(0.11)	0.84(0.14)	0.57(0.21)	0.48(0.21)

Note. N=36.

Table 3: Mean proportions correct and response times for letter feature tests, Experiment 2.

Mean proportions correct

	Distinct		Similar	
	3	5	3	5
Matching Target	0.91(.12)	0.89(.14)	0.93(.10)	0.90(.12)
Recombined Target	0.95(.09)	0.84(.18)	0.93(.12)	0.86(.14)
Lure	0.96(.07)	0.88(.16)	0.92(.11)	0.88(.14)

Median RTs

	Distinct		Similar	
	3	5	3	5
Matching Target	1813(645)	1992(580)	1747(425)	1883(584)
Recombined Target	1726(401)	1812(505)	1690(575)	1943(716)
Lure	1807(517)	1982(487)	1726(482)	2010(546)

Note. Incorrect responses and responses faster than 180 ms were excluded from the RT analysis. *N*=28.

Table 4: Proportions correct and response times for color feature tests, Experiment 2

Mean proportions correct

	Distinct		Similar	
	3	5	3	5
Matching Target	0.72(.20)	0.65(.22)	0.70(.17)	0.76(.22)
Recombined Target	0.72(.19)	0.63(.19)	0.72(.20)	0.76(.19)
Lure	0.87(.19)	0.80(.19)	0.57(.23)	0.36(.19)

Median RTs

	Distinct		Similar	
	3	5	3	5
Matching Target	1838(663)	1872(620)	1767(586)	1871(535)
Recombined Target	1733(561)	1949(1037)	1912(693)	1709(524)
Lure	1633(397)	1646(423)	1891(614)	2128(923)

Note. Incorrect responses and responses faster than 180 ms were excluded from the RT analysis. *N*=28.

Table 5: Mean number of correctly-recalled features (with standard deviations), Experiment 3

	Distinct		Similar	
	3	5	3	5
Letter features	2.83(0.25)	4.49(0.44)	2.88(0.19)	4.49(0.50)
Color features	2.24(0.34)	3.90(0.30)	1.69(0.38)	3.70(0.32)

Note. N=27.

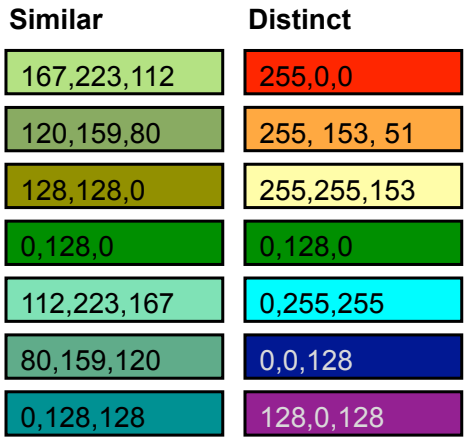

Figure 1. The similar and distinct color sets of Experiment 1, with RGB coordinates.

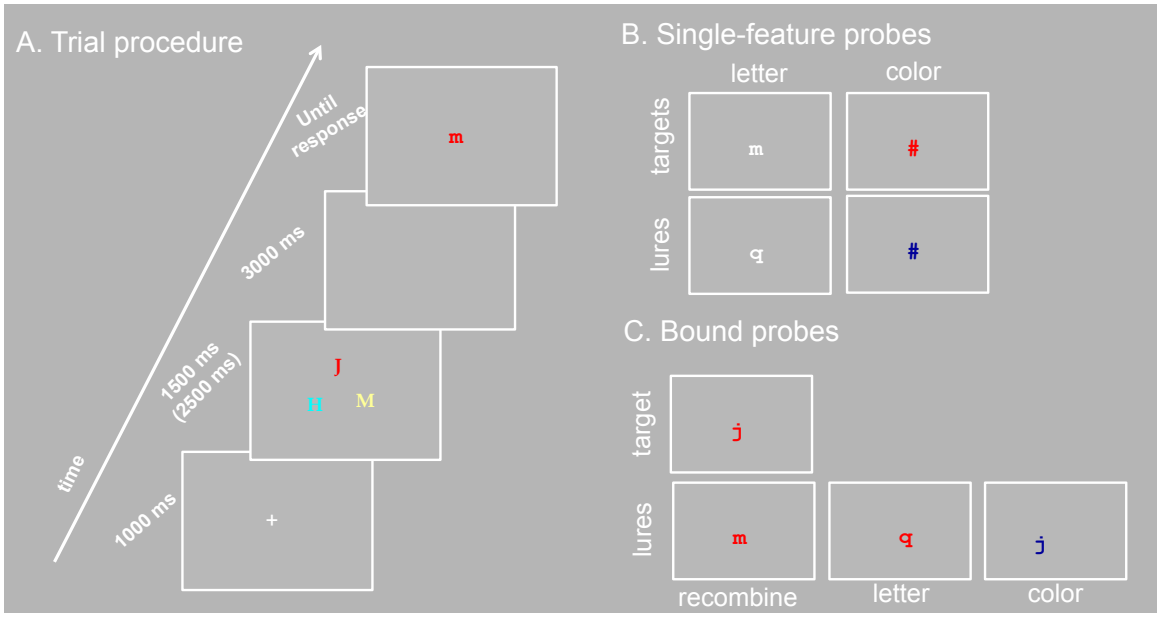

Figure 2. A) Experiment 1 trial procedure. In this example, the test is a recombination lure, and the correct response would be "no", indicating that there was no red m in the study array. B) Examples of possible single-feature probes. C) Examples of possible bound probes. In all blocks, 50% of trials were targets and 50% were lures. Images are not drawn to scale.

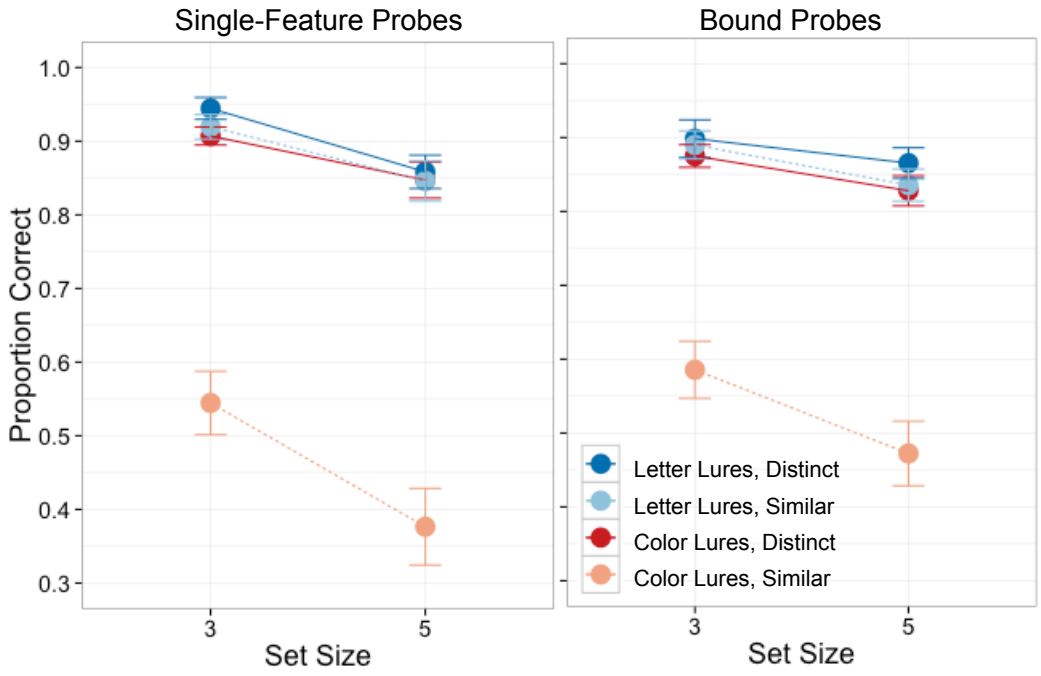

Figure 3. Accuracy on color and letter lure trials, Experiment 1. Error bars are within-subject standard errors of the mean with the Cousineau-Morey correction applied (Cousineau, 2005; Morey, 2008).

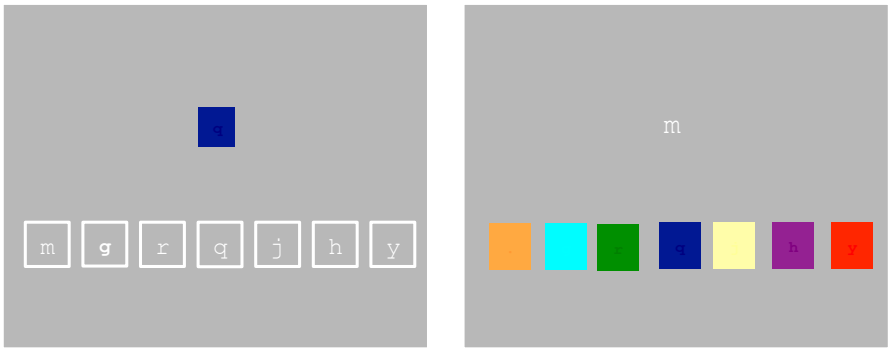

Figure 4. Choice recognition binding tests of Experiment 4. On the left is an example in which color is the eliciting feature; here, participants must choose the letter that was presented in the given color on that trial. In the example on the right, participants would choose the color that was presented with the given letter on that trial.

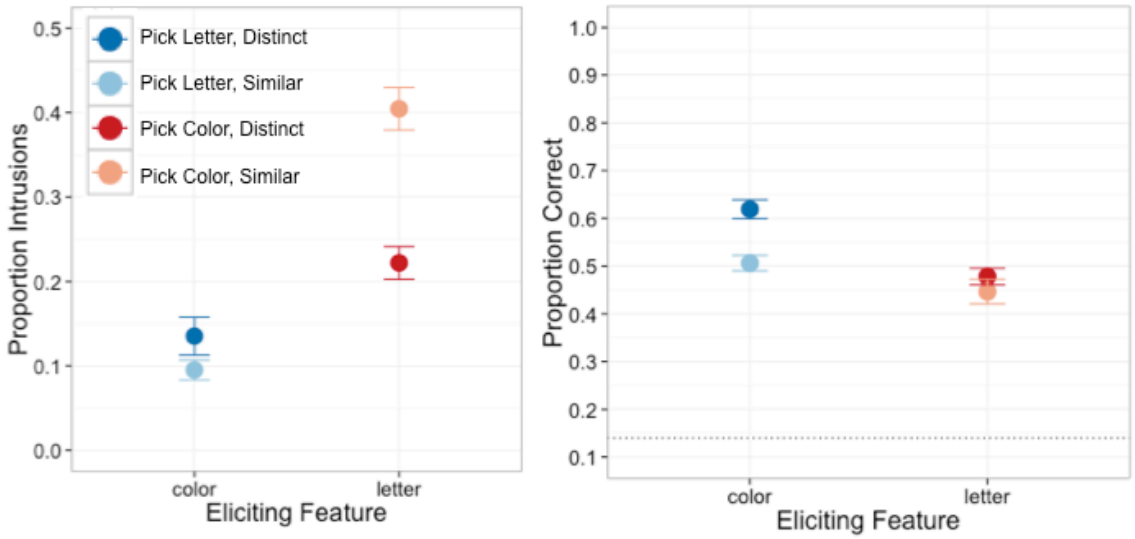

Figure 5. Mean proportions intrusion errors (left) and proportions correct excluding intrusion errors (right), by eliciting feature and color similarity. The dotted horizontal line in the right panel represents chance performance (i.e., 1/7). Error bars are within-subject standard errors of the mean calculated with the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).